Data Analysis Tools

**Assignment – Week 4**

**Testing a Potential Moderator**

**By mapolarbear@gmail.com**

The assignment deals with testing a potential moderator. When testing a potential moderator, we are asking the question whether there is an association between two constructs for different subgroups within the sample.

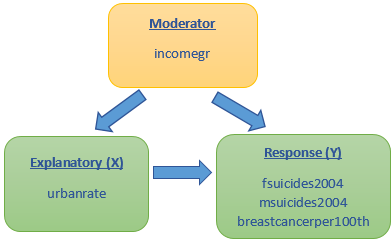
**Research question:** If moderator variable **incomegr** has association/effect on 2004 suicide rates of females and males. In case if there is association between moderator variable and research variables is it similar or different for female and male suicide rates.

**Variables:**

**fsuicides2004, msuicides2004, urbanrate, breastcancerper100th -** All variables are QUANTITATIVE.

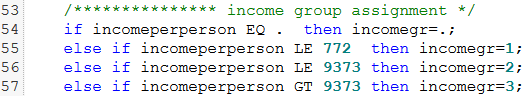
**Incomegr** – Moderator variable/construct, QUANTITATIVE, created form **incomeperpreson** variable.

General moderation concept schema:

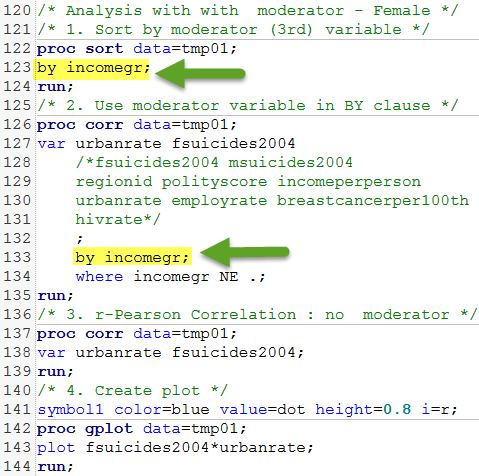


**Program code:**

1. Create **incomegr** variable (**moderator**)



1. After incomegr variable added to SAS dataset, then, before execution of any test like **ANOVA, CHI-Square or Perason Correlation**, we have to sort dataset by moderator variable. The sort step must include **BY** clause followed by moderator variable.
2. After dataset sorted, add moderator variable using BY close to our test. In my case, as long as my base study involves quantitative variable/indicators **fsuicides2004, msuicides2004, urbanrate, breastcancerper100th,** the **r-Pearson Correlation coefficient** (**PROC CORR**) was used. In addition, I unclouded execution of Pearson Correlation (PROC CORR) without moderator and for graphical representation of the same result, PLOT (PROC GPLOT) to better understand and interpret results, and to see full picture of how explanatory and response variable correlated. Here is the code:



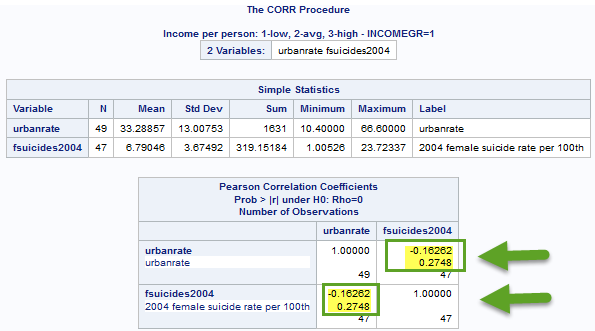
**Output Results and interpretations**

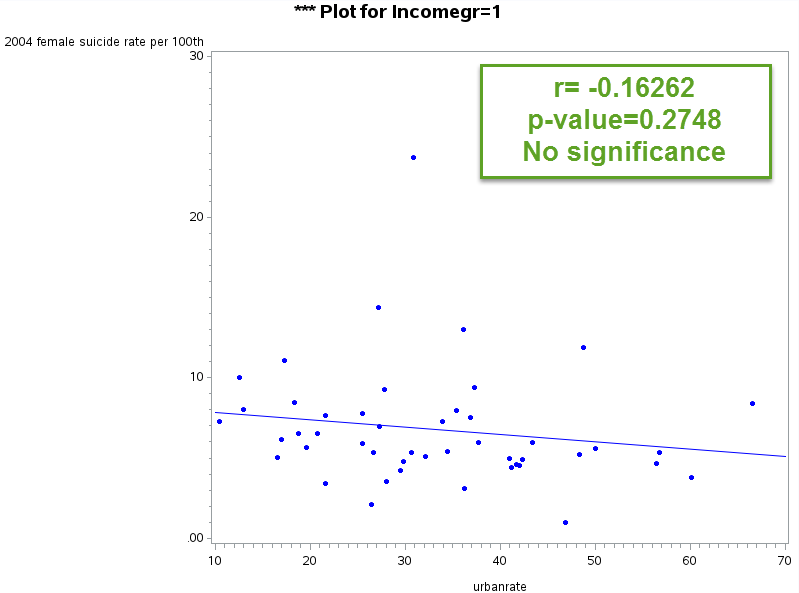
1. Analysis with MODERATOR variable **incomegr:**

There are 3(three) groups in incomegr: 1 – low income group (below 772 income per person), 2- average income group (between 772 and 9373 income per person) and 3 – high income group (values higher than 9373 income per person).

**1 – Low income group** (below 772 income per person)

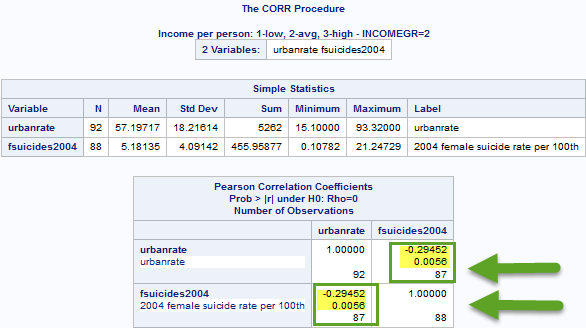
Result for group 1 shows **weak negative linear correlation** between **urbanrate and fsuicides2004** **with r-Pearson Coefficient value or -0.16262** is close to 0 (zero) and **p-value of 0.2748**, which is greater 0.05 and **we can conclude that the association between urbanrate and female suicides is not significant in countries with low income.**

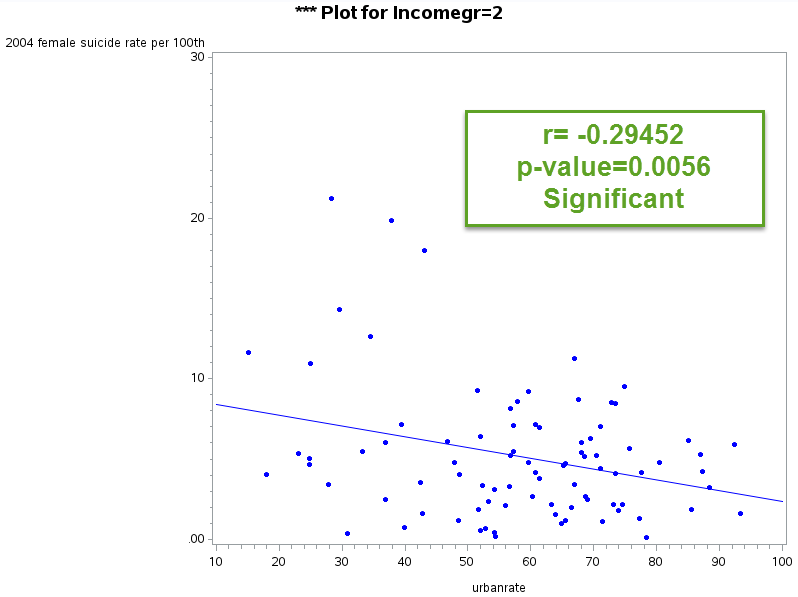




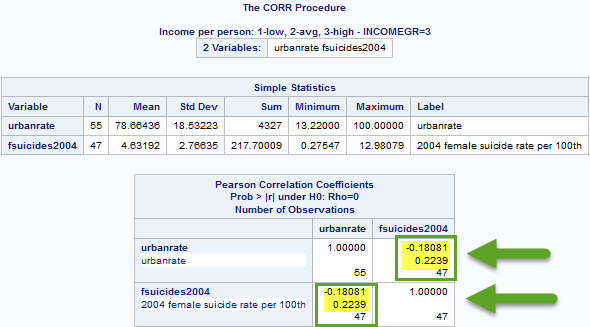
**2 – Average income group** (between 772 and 9373 income per person)

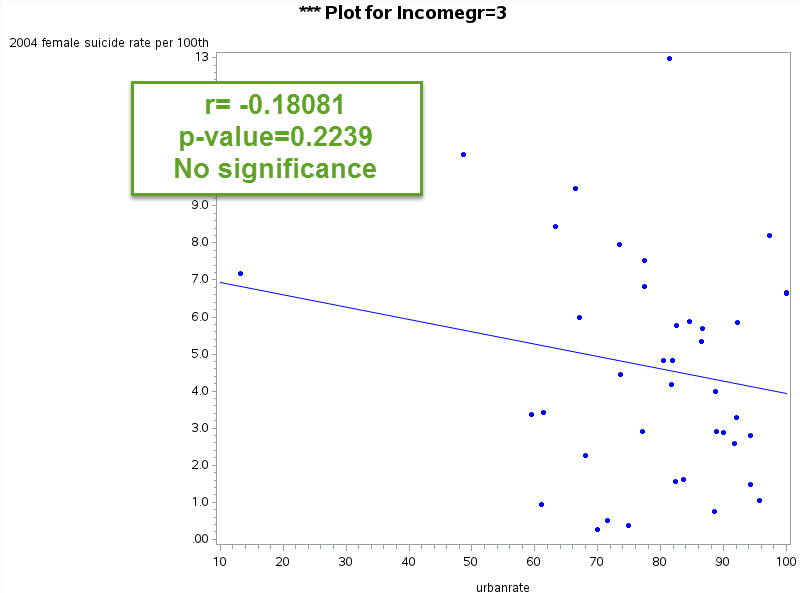
Result for group 2 shows **strong negative linear correlation** between **urbanrate and fsuicides2004** **with r-Pearson Coefficient value or -0.29452** is close to 0 (zero) and **p-value of 0.0056**, which is less 0.05 and **we can conclude that the association between urbanrate and female suicides is significant in countries with average income.**





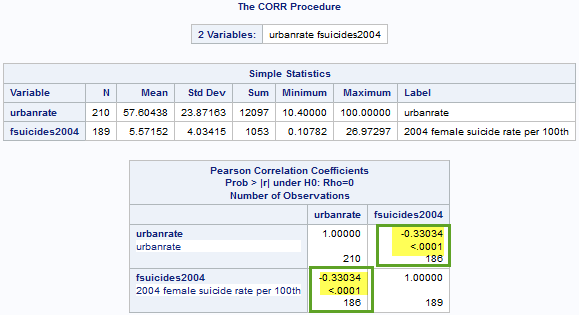
**3 – High income group (values higher than 9373 income per person)**. **weak negative linear correlation** between **urbanrate and fsuicides2004** **with r-Pearson Coefficient value or -0.18081** is close to 0 (zero) and **p-value of 0.2239**, which is greater 0.05 and **we can conclude that the association between urbanrate and female suicides is not significant in countries with high income.**

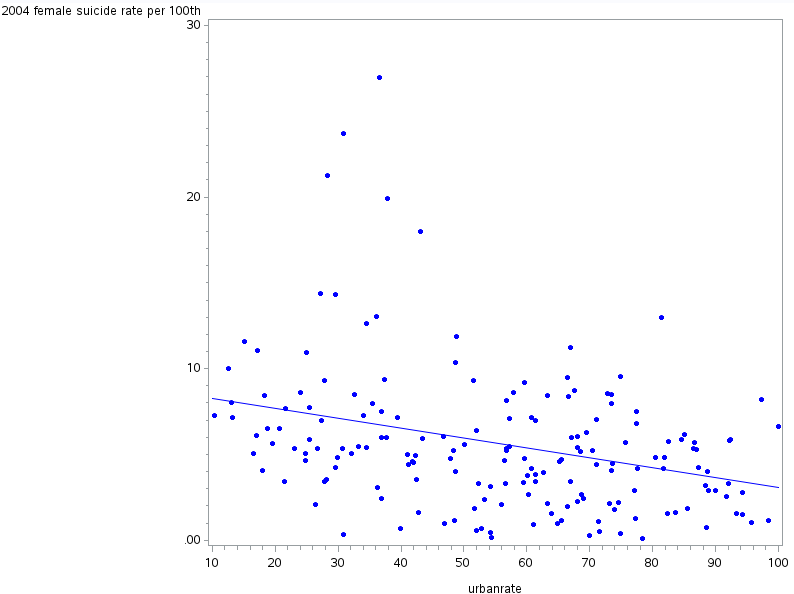




In addition to above study I ran similar test against **msuicides2004** and **result didn’t show significant association between urban rates and male suicides in any income group rate**. This is very different from female suicides study.

In addition to above study I ran Linear correlation coefficient test between fsuicides2004 and urbanrate, just to better understanding how variables are correlated and result shows very strong linear correlation. Based on results of PROC CORR, variables have strong **negative linear correlation of -0.33034 and p-value or < 0.0001 is significant**. The corresponding PLOT also proves strong relationship between variables. We may also roughly conclude that country with higher urbane rate have lower female suicide rating. But of cause, there are many other factors affecting suicide rating, urbanization is just one of them





\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* End of week 3 assignment \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

<http://coursera-sas-tools-a-week-1.tumblr.com/post/152079337232/data-analysis-tools-assignment-week-1-running>

<http://coursera-sas-dm-week4.tumblr.com/post/151798525721/data-management-and-visualization-assignment>